Machine Learning MC321
Lab Assignment 4

February 15, 2025

Implement the kNN classifier and evaluate the performance on "Social Network Ads.csv" which is a categorical dataset to determine whether a user purchased a product or not by using three features to determine user's decision.

- Visualize the data by 3D plotting features using different colors for label 0 and 1.
- Compare your kNN classifier's implementation with the available library function. Use 90% data points for training and the remaining 10% for testing the accuracy of classification.
- Using the confusion matrix find accuracy, precision, F1 score and recall for different values of k.
- Use the K-fold cross-validation technique (assume $K = 5$) and plot the average accuracy as a function of k neighbours. Which value of k leads to the best accuracy?
- Effect of k: Show the decision boundaries considering two features at a time for different values of k. Which values of k result in smoother boundaries?
- Regression: Implement kNN for regression. Consider the first two features as the input and the third feature as the output variable and assume 60:40 train-test split. Report the MSE for different values of k on the test data.
- "Use bootstrapping to generate multiple training datasets. Train a K-Nearest Neighbors (K-NN) model on each bootstrap sample and estimate the prediction error. Compare the average bootstrapped error with the error from the original dataset."